Introduction of AI (CS103)- 08 Machine Learning Algorithms

Jimmy Liu 刘江

2023-10-10

Latest AI development: GPT4- Turbo

All Collections › OpenAI API › GPT-4 Turbo

GPT-4 Turbo

Our latest model

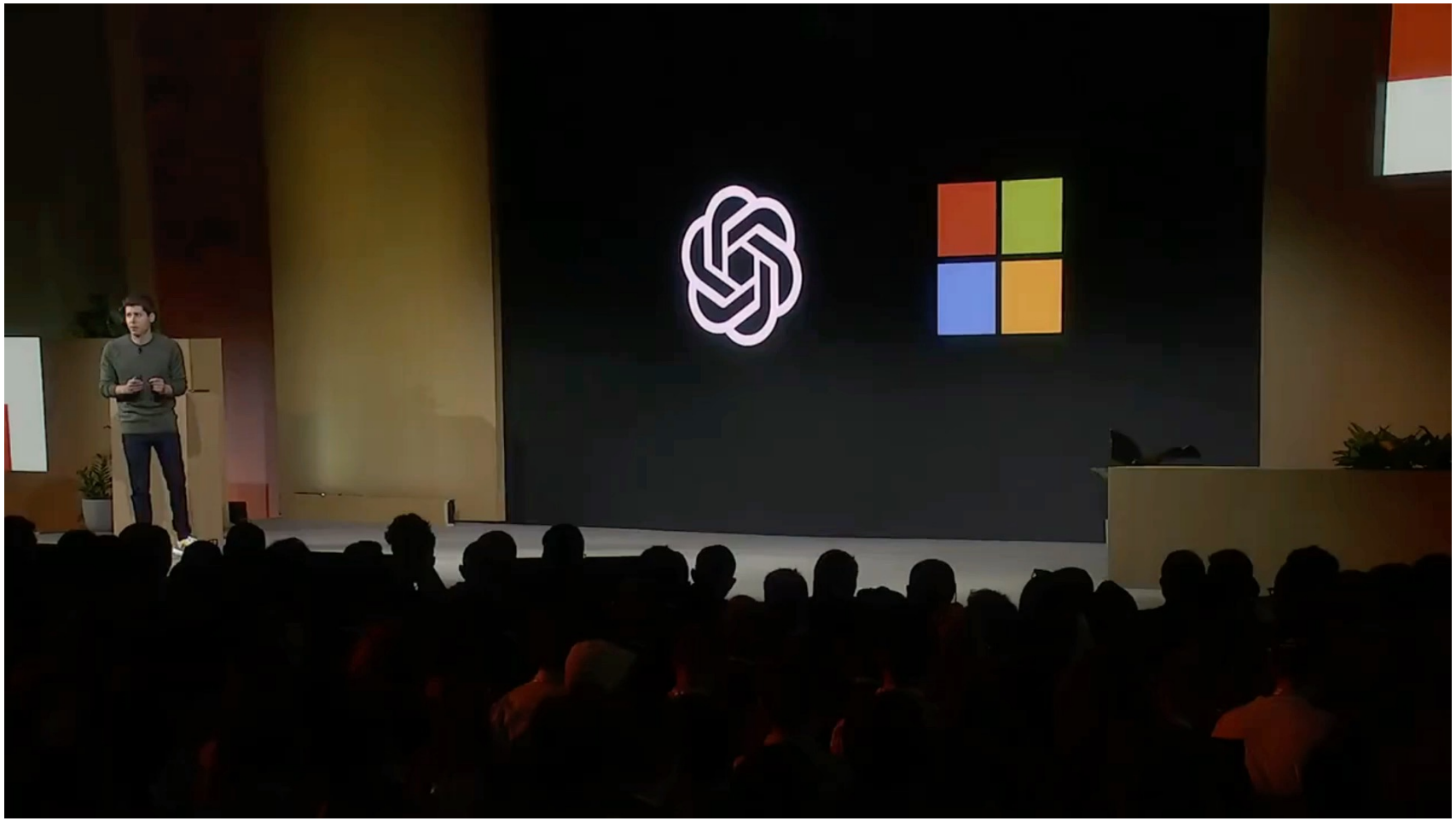
Written by Raymond Niles
Updated this week

What is it?

GPT-4 Turbo is our latest generation model. It's more capable, has an updated knowledge cutoff of April 2023 and introduces a 128k context window (the equivalent of 300 pages of text in a single prompt). The model is also 3X cheaper for input tokens and 2X cheaper for output tokens compared to the original GPT-4 model. The maximum number of output tokens for this model is 4096.

GPT-4 Turbo 6 Improvements

GPT-4 Turbo’s Developing Platform

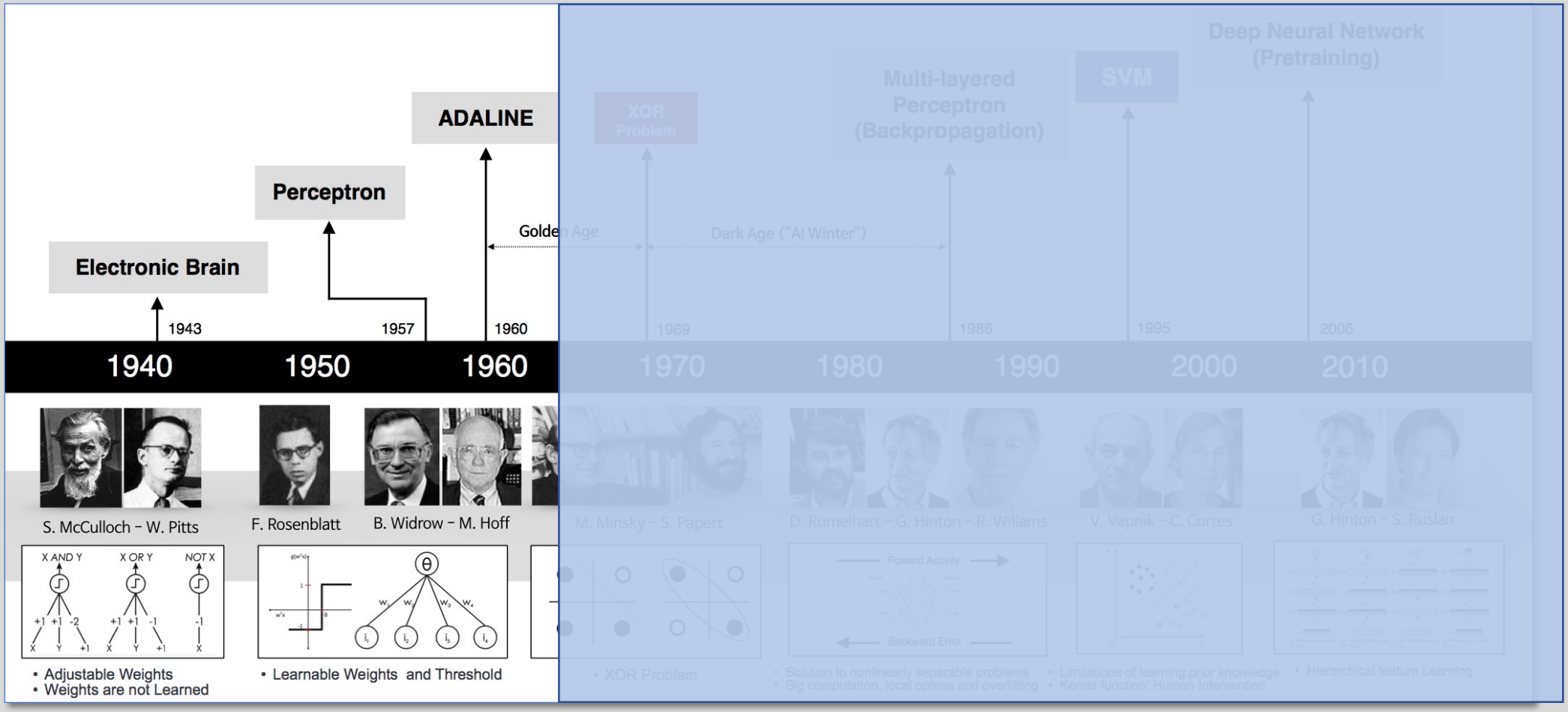


Self-Defined GPT

Lecture 8

1. Reviews of Lecture 7 and Mid-Term Test
2. ADALINE and XOR Logic
3. Support Vector Machine (SVM)
4. Decision Tree

AI Development Milestones-ADALINE

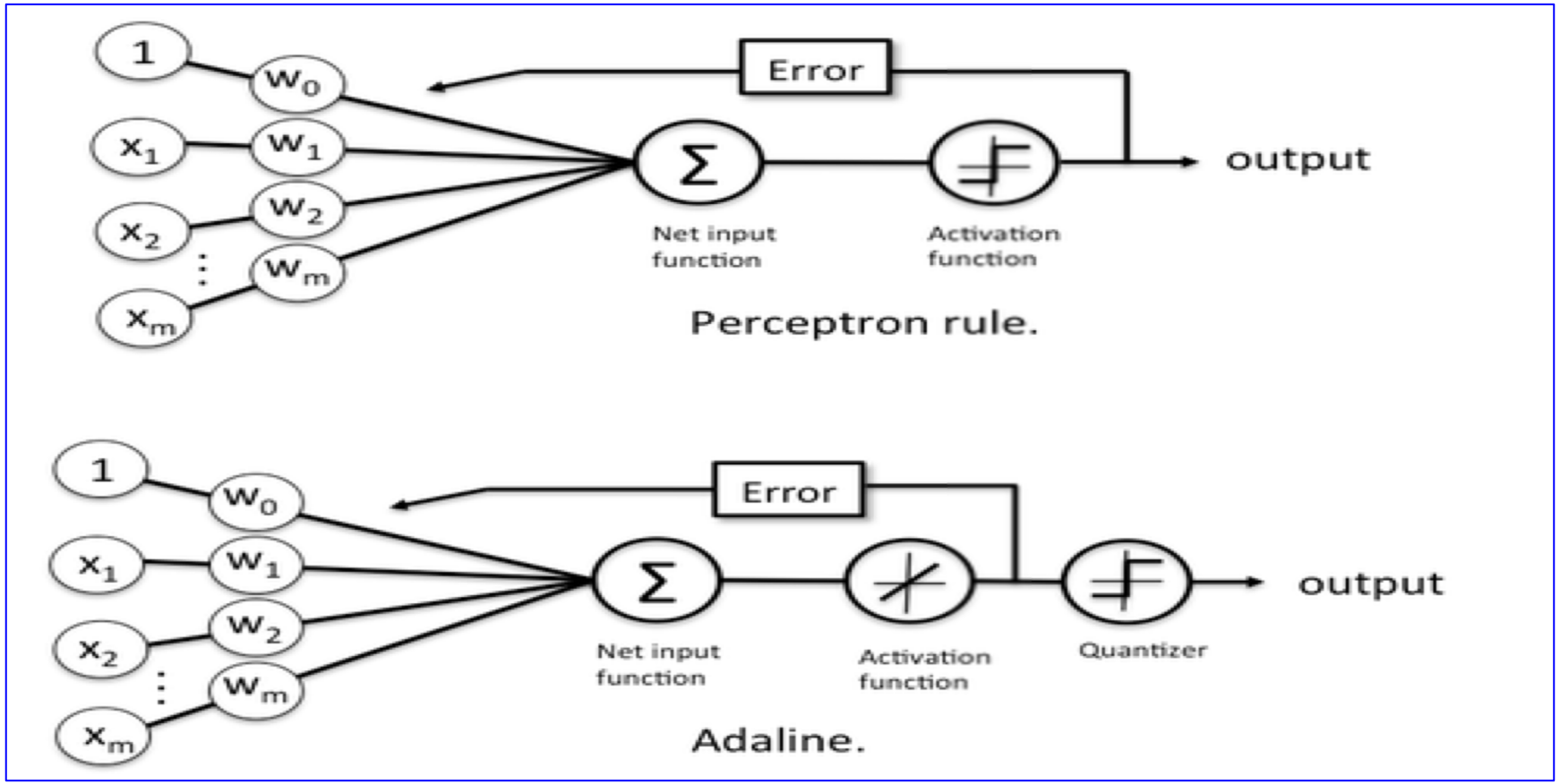


ADALINE (Adaptive Linear Neuron)

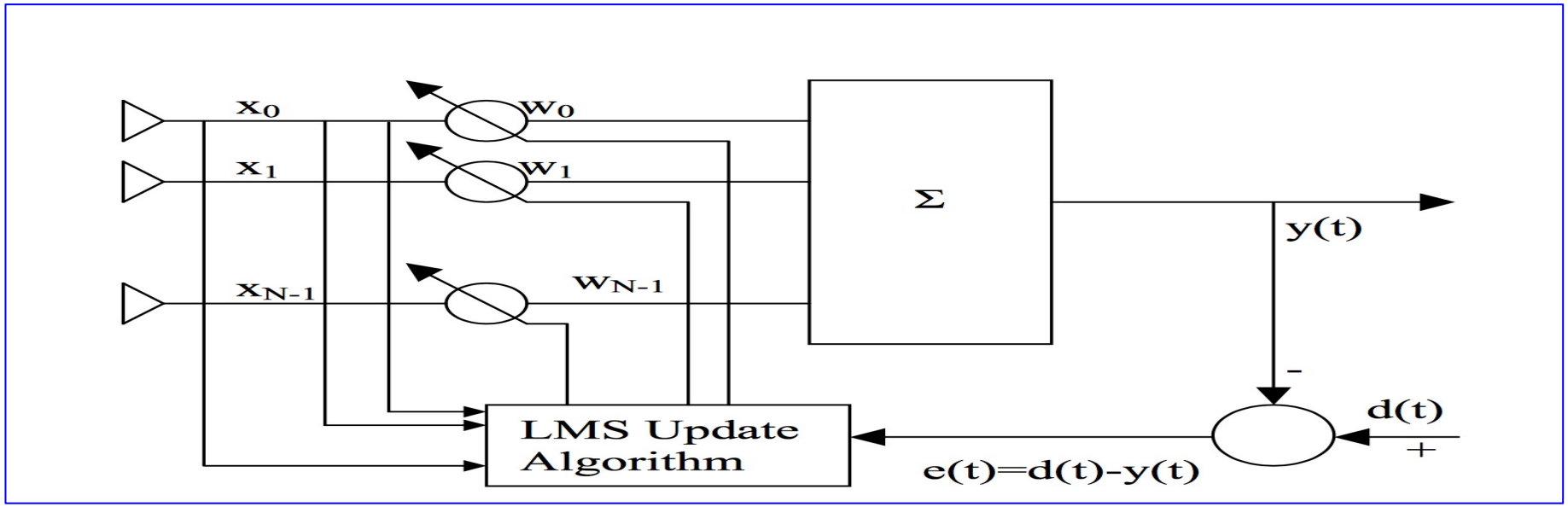
ADALINE is a single layer neural network based on MSE and DELTA Supervised learning

ADALINE is proposed in 1960 by Stanford ProfessorBernard Widrow and his first PhD student Ted Hoff

ADALINE


1
w_0
x_1
w_1
w_2
x_2
w_m
x_m
Error
Σ
output
Net input function
Activation function
Perceptron rule.
1
w_0
x_1
w_1
w_2
x_2
w_m
x_m
Error
Σ
output
Net input function
Activation function
Quantizer
Adaline.

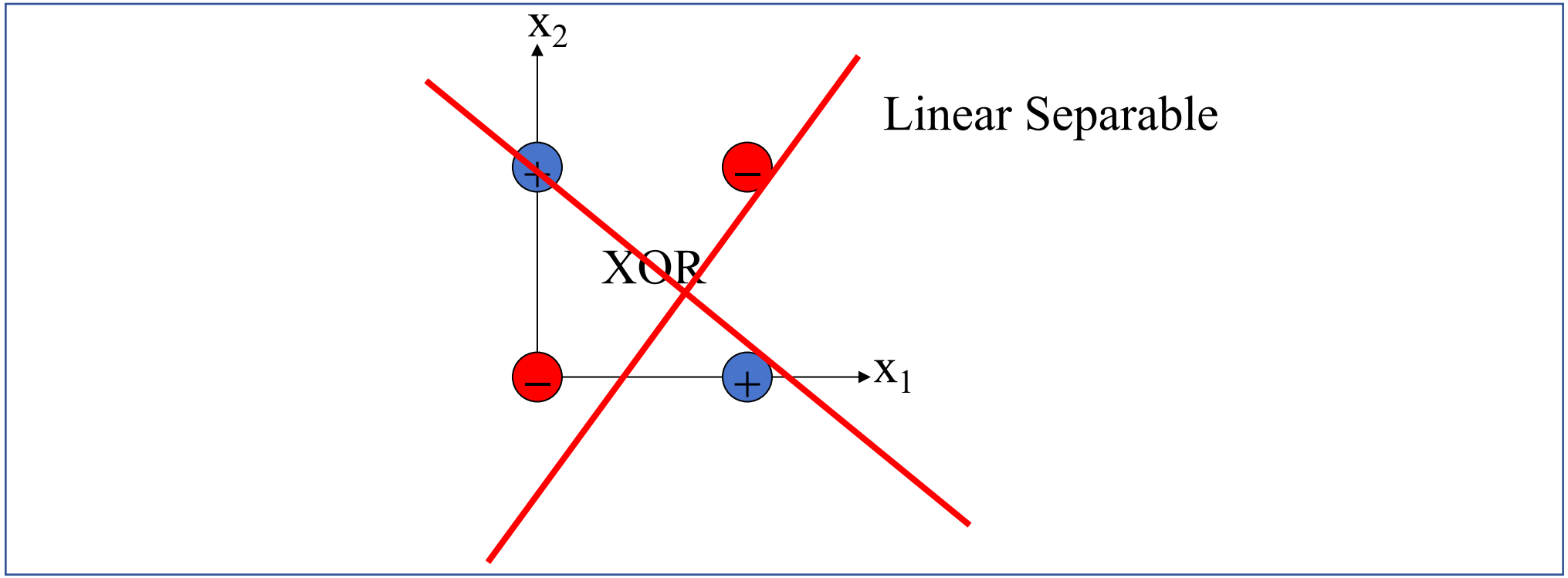
Widrow Hoff Delta Learning Rule



$$\Delta w_i = \eta(t-o)x_i$$

** B.Widrow and M.E.Hoff, "Adaptive switching circuits," Proc. Of WESCON Conv. Rec., part 4, pp.96-140, 1960*

Perceptron's XOR Logic Challenge



Minsky & Papert (1969)

Bad News: Perceptron can only handle linear separable problem

Q1: How to Prove Perceptron Can not Implement XOR Logic

AI Winter

- Perceptron can not implement XOR Logic
  - 1950s, English to Russian Translation

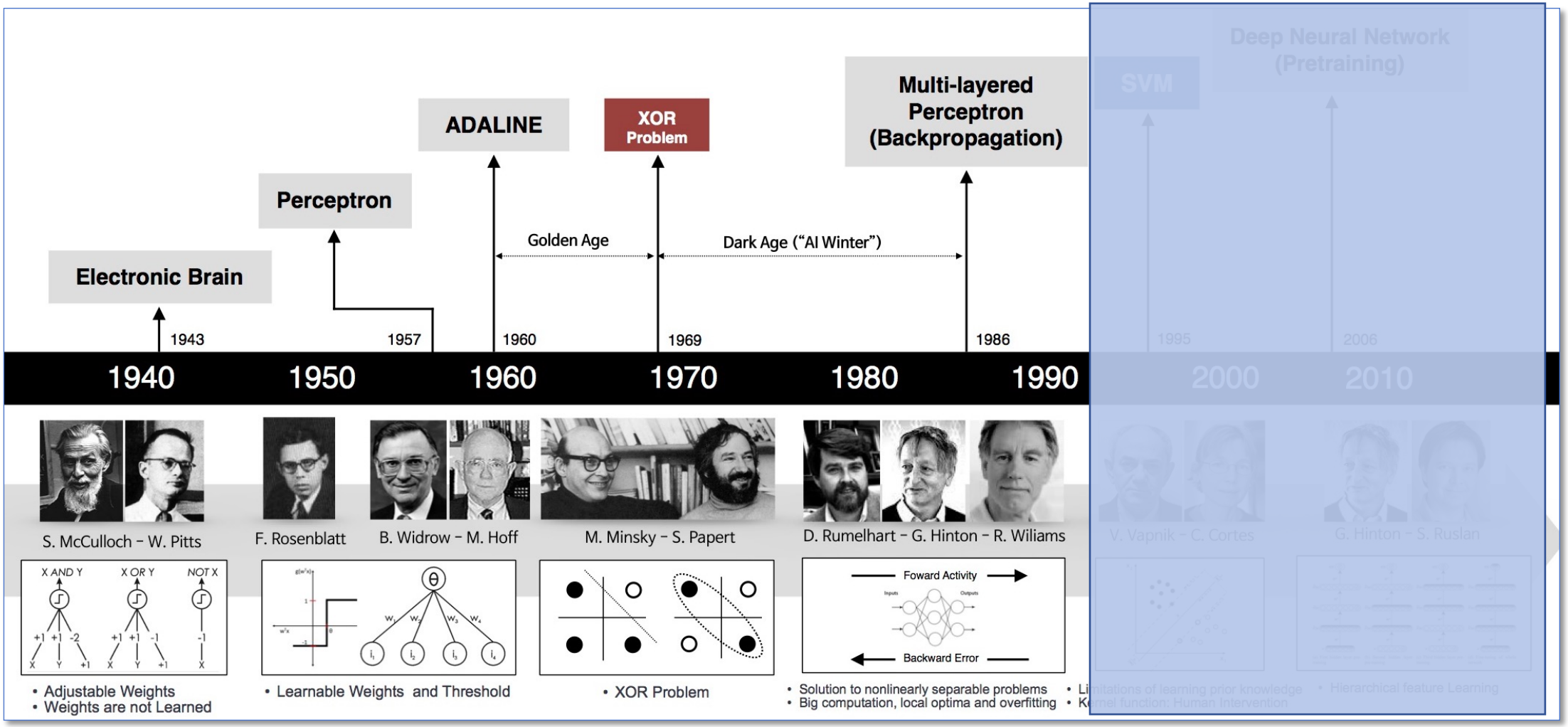
    “The spirit is willing but the flesh is weak”

    is translated as:

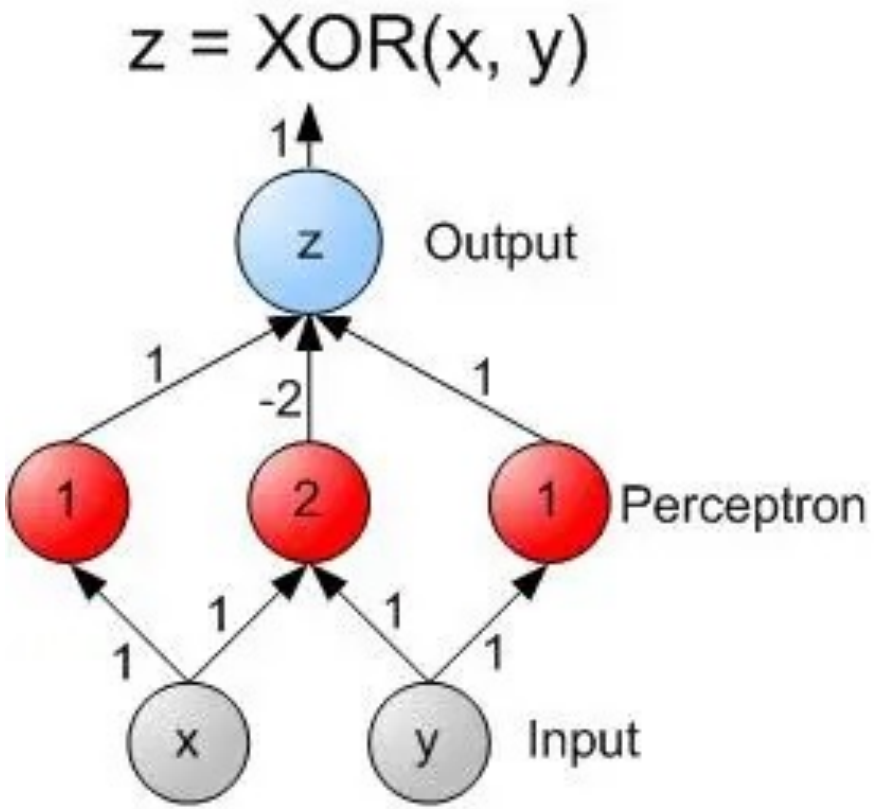
    The vodka is good but the meat is rotten”
- AI system can not incorporate knowledge
- Funding greatly reduced

Q2: Any Solution To Overcome the Perceptron's Limitation

AI Development Milestones
- Multi-Layer Perceptron

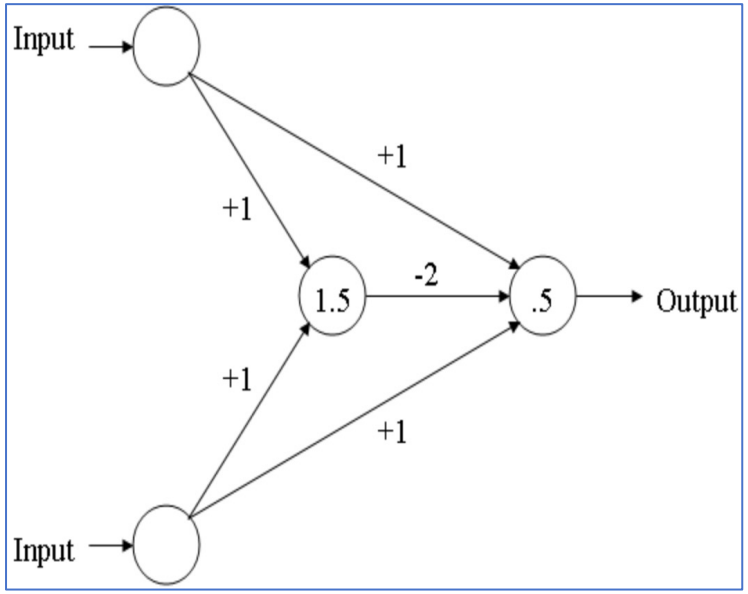


Multi-Layer Perceptron



Homework 8-1 ： Prove The Following MLP Can Implement XOR Logic

1

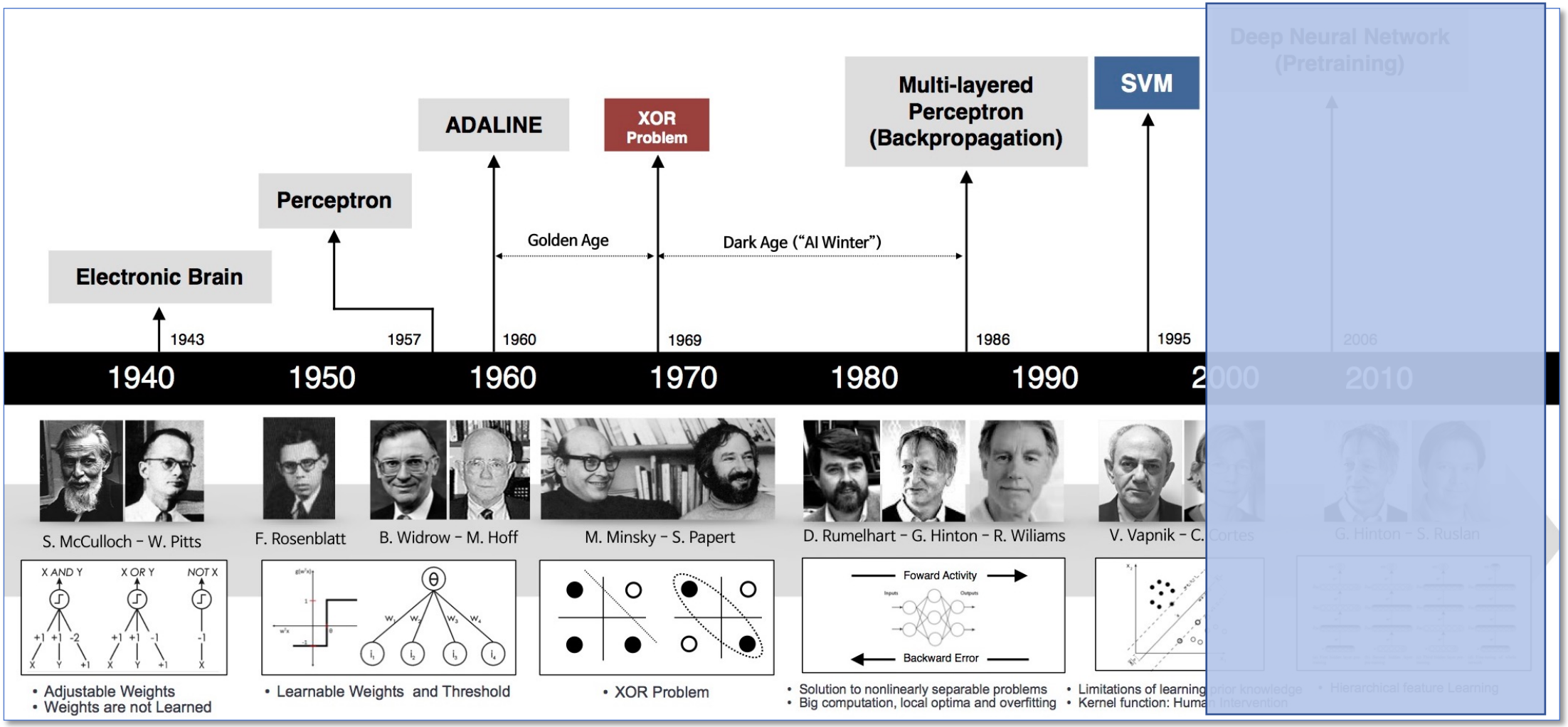


Input		Output
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Lecture 8

1. Reviews of Lecture 7 and Mid-Term Test
2. ADALINE and XOR Logic
3. Support Vector Machine (SVM)
4. Decision Tree

AI Development Milestones - SVM



Support Vector Machines

Support vector machine(SVM) was proposed by Vladimir N. Vapnik and Alexey Ya Chervonenkis in 1963 for the purpose of solving 2-class problem.

Vladimir Vapnik NEC,
U Colombia,
Royal Holloway London

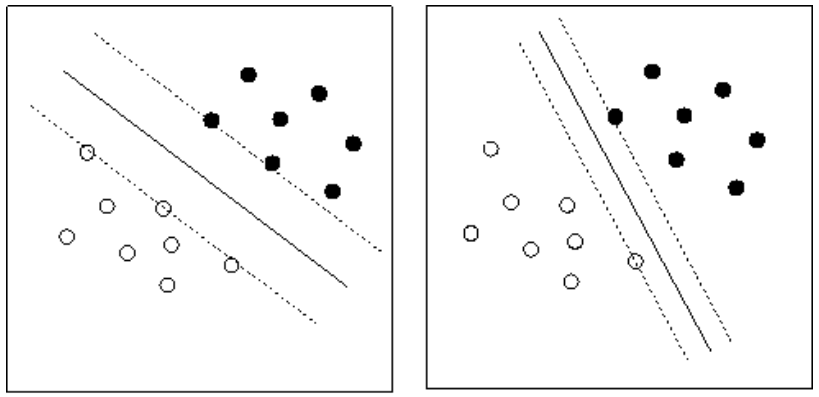
Support Vector Machines

SVM has become a general machine learning algorithm and has been successfully applied in many fields.

For example, using SVM to process mail information, you can determine whether an email is spam.

Linear Separatable Problem

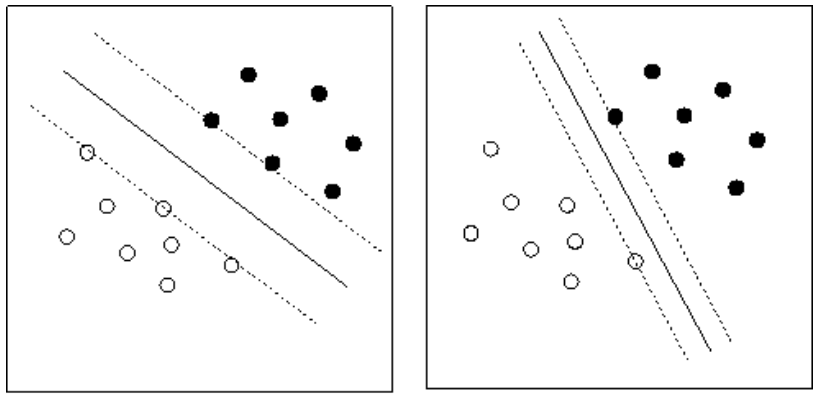
In two-dimensional coordinate space, **linear separatable problem** is means that we can use one straight line to separate two different categories completely



Linear Separatable Problem

Mathematical definition：

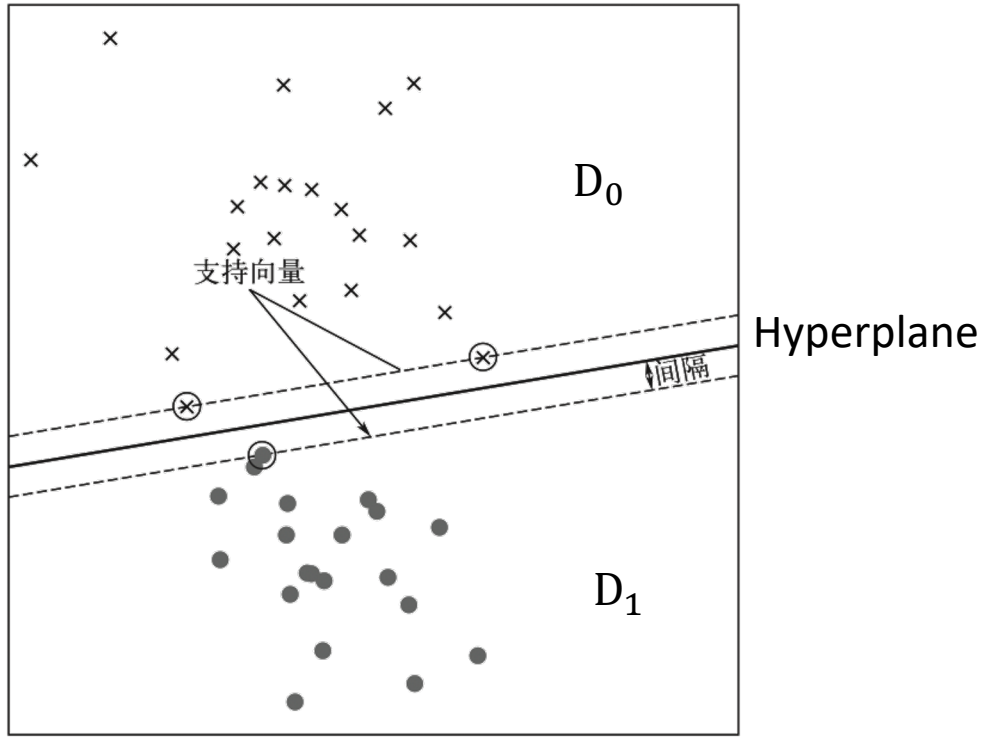
Assume D_0 and D_1 are two sets of points in Euclidean space , if there is an n-dimensional vector w and a real number b, such that all points x_i in D_0 that satisfy $\omega x_i + \mathrm{b} > 0$, and all points x_j in D_1 that satisfy $\omega x_j + \mathrm{b} < 0$, we can declare that D_0 and D_1 are linear separable.



Maximum Margin Hyperplanes

A Maximum Margin Hyperplane is a plane that separates two sets of points (D_0 and D_1) in such a way that the distance between the closest points in the sets to the hyperplane is maximized.

- Importance:
- - Better generalization
- - Minimizes overfitting

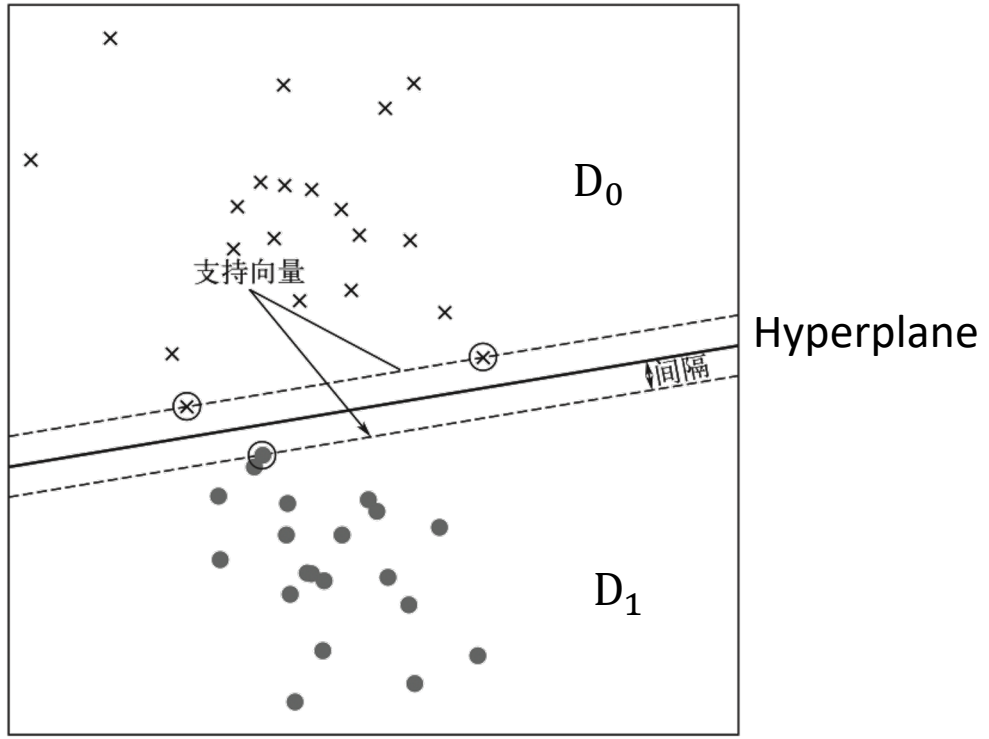


What are Support Vectors?

Support Vectors are the points in sets D_0 and D_1 that are closest to the Maximum Margin Hyperplane.

Importance:

- Define the Maximum Margin Hyperplane
- Critical for model's performance



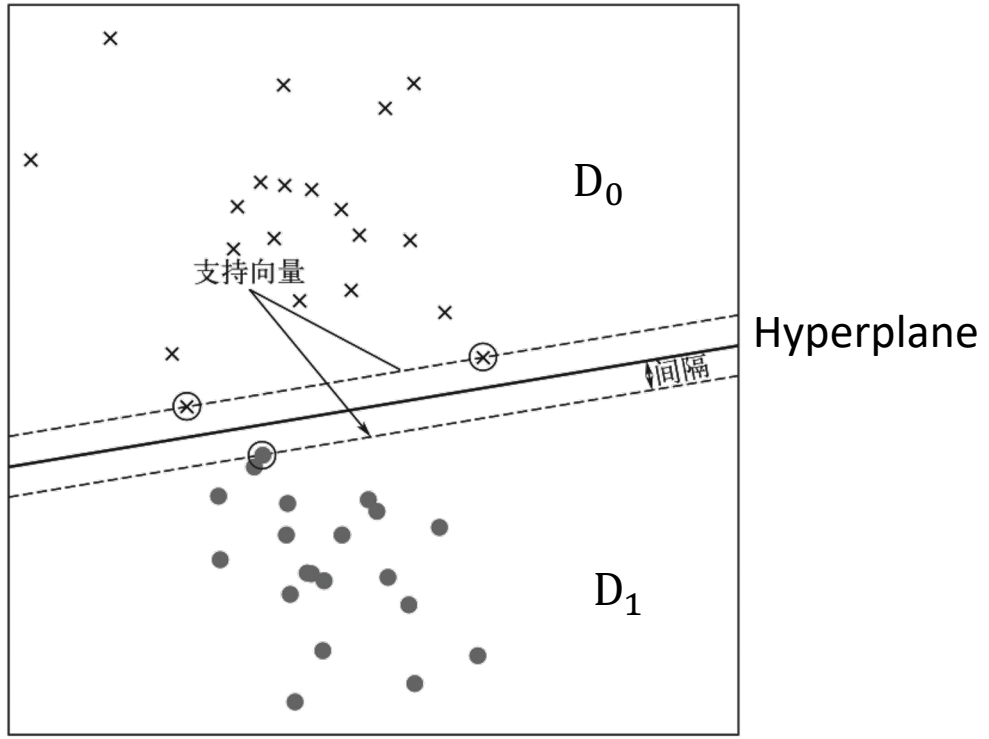
What is the Margin?

The Margin is the distance between the Support Vectors and the Maximum Margin Hyperplane.

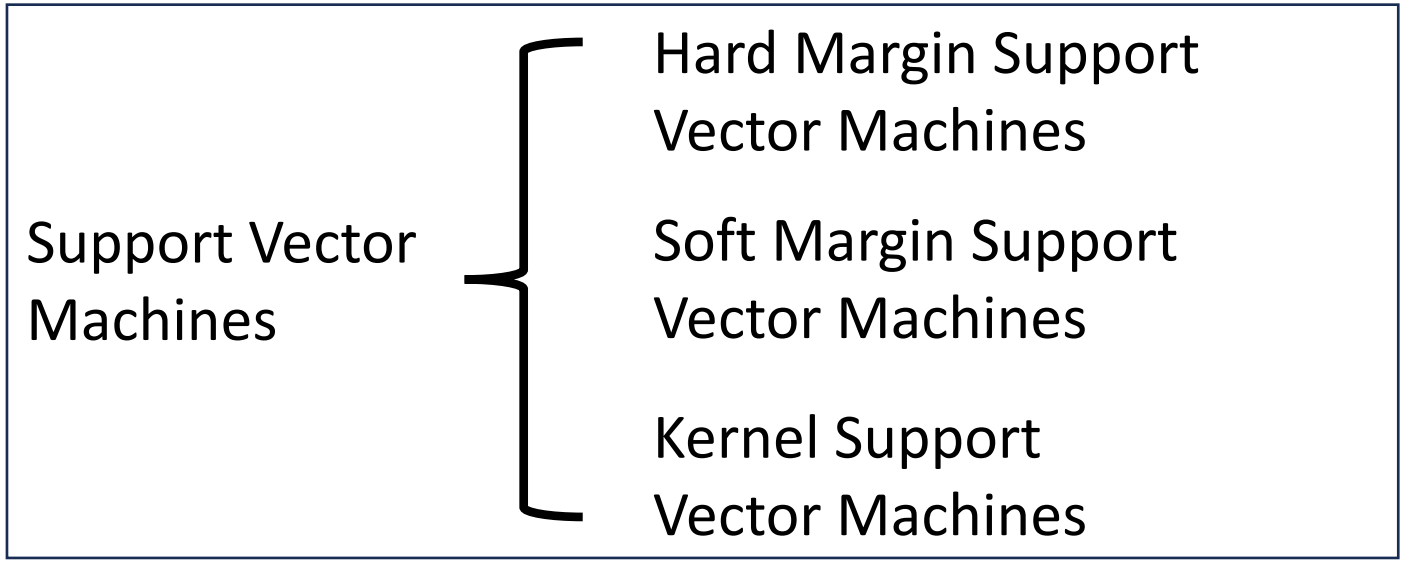
Significance:

- Larger margin usually leads to better generalization

- Smaller margin can cause overfitting



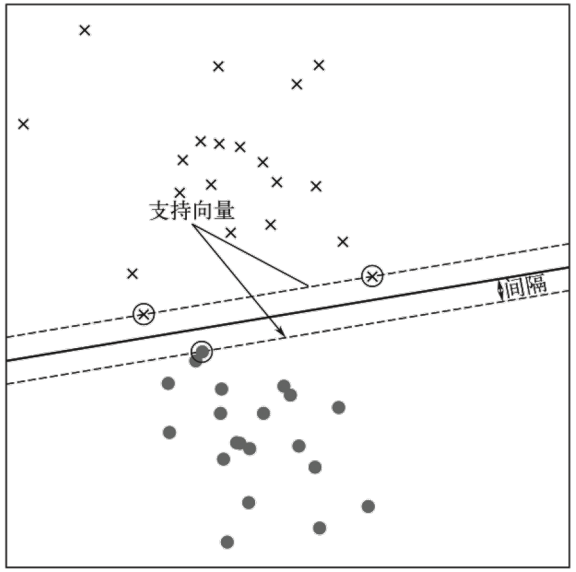
Classification for Support Vector Machines



What is a Hard Margin SVM

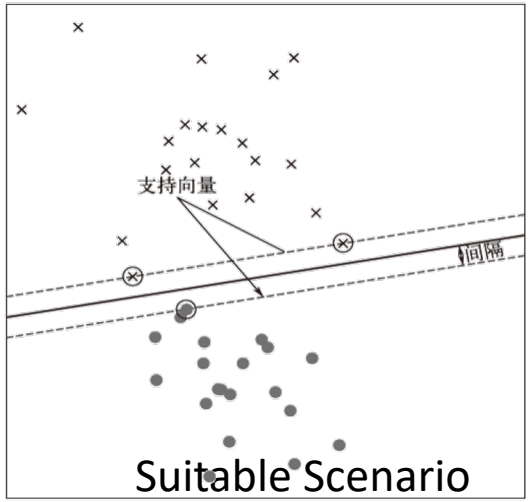
Hard Margin Support Vector Machine (SVM) is a type of SVM that strictly separates data points of different classes without any overlap or misclassification.

It is ideal for linearly separable data.

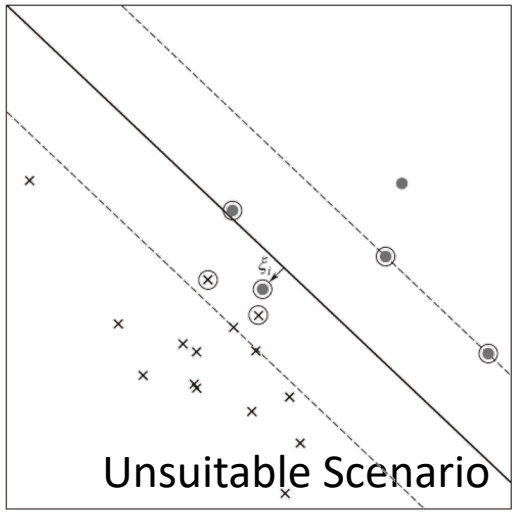


Hard Margin SVM

- Hard margin support vector machines ensure perfect classification separation, but they are sensitive to outliers and only apply to linearly separable data.
- The goal of hard spaced support vector machines is to maximize margins while ensuring correct classification.



Suitable Scenario


Unsuitable Scenario

Mathematical Formulation(1)

- The objective function for Hard Margin SVM is:

$$\max \frac{1}{||w||} miny(w^T x + b)$$

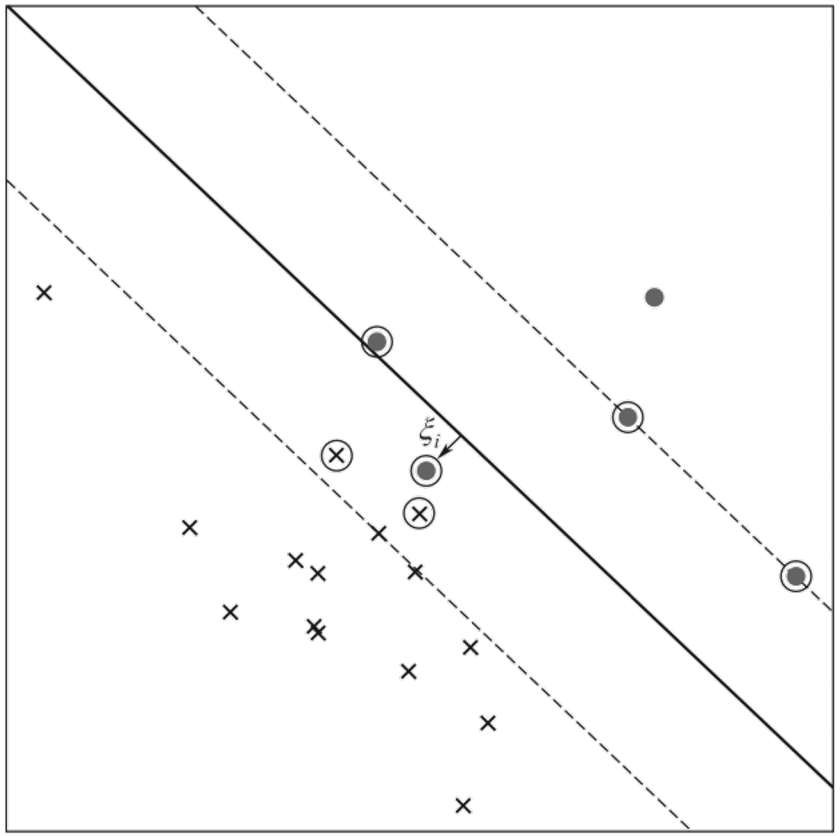
- Subject to:

$$y \frac{(w^T x + b)}{||w||} \geq 1$$

What is Soft Margin SVM?

- Soft Margin Support Vector Machine (SVM), on the other hand, is a variant of SVM that allows some level of misclassification.
- It is more robust to outliers and is applicable to non-linearly separable data.
- The objective function is:

$$\max \frac{1}{\| w \|} \min y\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b\right) + \sum_{i=1}^{m} \xi_i$$



Mathematical Formulation(2)

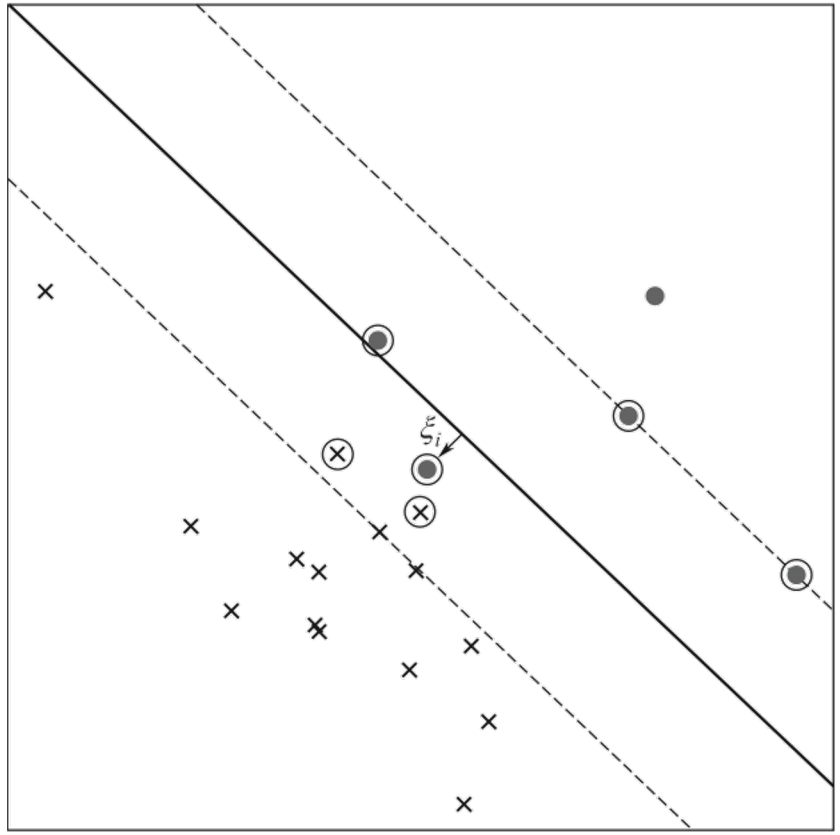
- In Soft Margin SVM, slack variables ξ_i are introduced to allow misclassification.

- The objective function becomes:

$$\max \frac{1}{\| w \|} \min y\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b\right) + \sum_{i=1}^{m} \xi_i$$

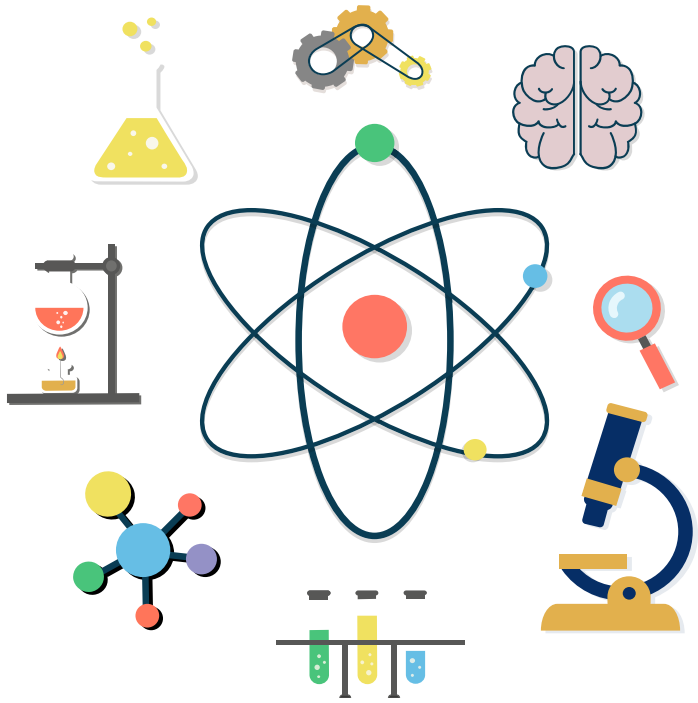
Why Soft Margin?

- Soft Margin SVM is often preferable when the data is noisy or non-linearly separable.
- It introduces a trade-off between maximizing the margin and minimizing misclassification.



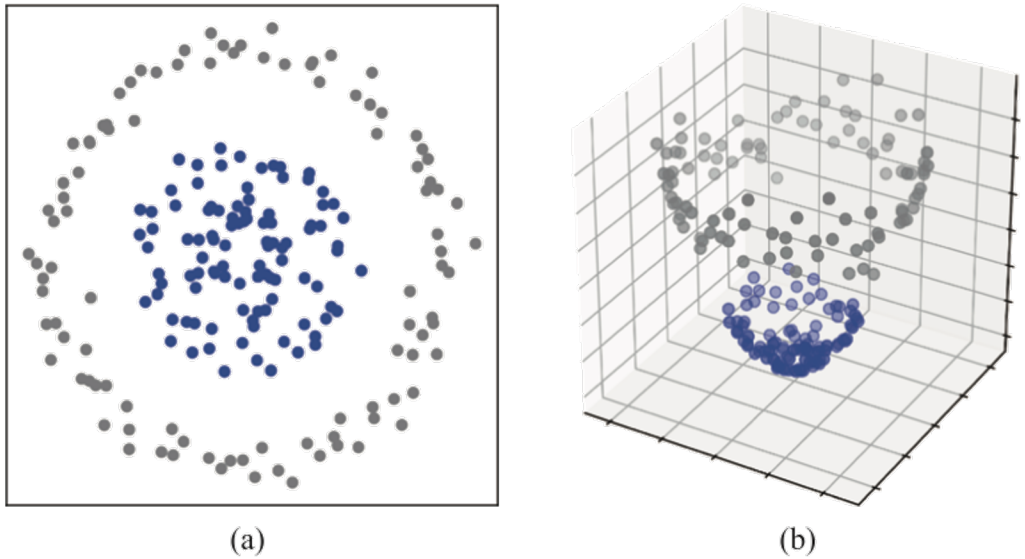
Hard Margin vs Soft Margin

- Hard Margin SVM is strict about data separation whereas Soft Margin SVM allows some misclassification.
- Soft Margin is generally more robust to outliers and applicable to non-linearly separable data.



What are Kernel Functions?

- Kernel functions are used in SVM to transform the input data into a higher-dimensional space. This makes it easier to find a hyperplane that separates the data. Common types include linear, polynomial, and radial basis function (RBF) kernels.



(a) (b)

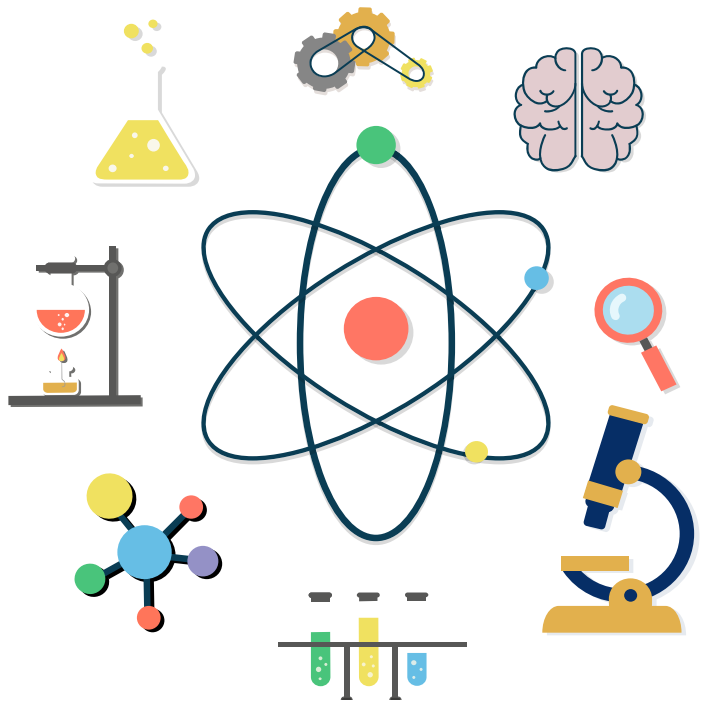
Types and Applications of Kernel Functions

1. **Linear Kernel: Suitable for linearly separable data.**

$$k(x, x') = x^T x'$$

2. Polynomial Kernel: Useful for data that has a polynomial degree of separation.

$$k(x, x') = (x^T x' + c)^M$$

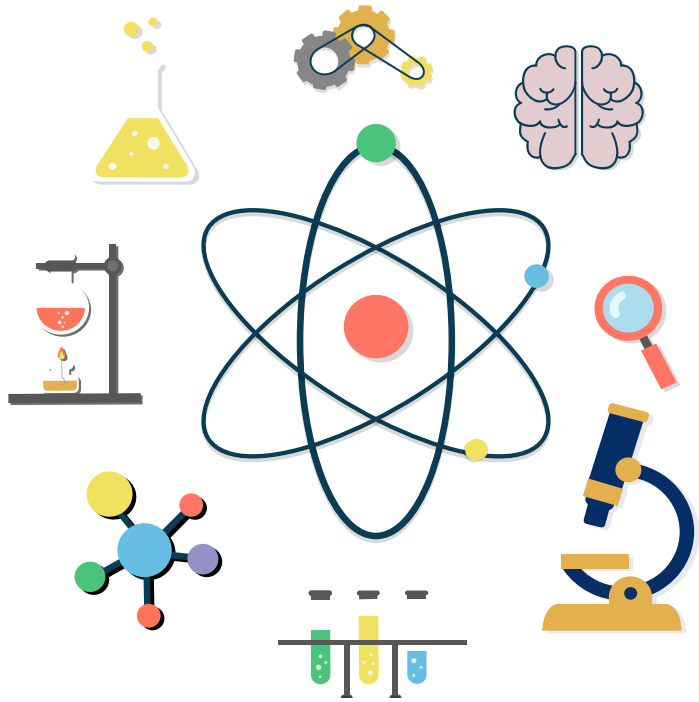


Types and Applications of Kernel Functions

3. RBF Kernel: Effective for non-linearly separable data.

$$k(x, x') = \exp(\frac{-||x - x'||^2}{2\sigma^2})$$

Applications range from text classification to image recognition.



Advantages and Disadvantages of SVM

Advantages:

1. Effective in high-dimensional spaces.
2. Memory efficient.
3. Versatile: Different Kernel functions can be specified.

Disadvantages:

1. Not suitable for large datasets due to high training time.
2. Less effective on noisier datasets with overlapping classes.

Example Comparison

Table: Distance from each sample in the myopia dataset to the hyperplane

Method	w	b	x	d
SVM	(–0.0150, –0.1445, 0.0780)	–2.6062	(3.7, –10.5, 27.49)	6.07
			(4.8, 0.13, 21.75)	6.07
			(4.9, 1.25, 22.31)	6.79
			(3.8, –6.38, 24.71)	1.13
			(5.0, –0.38, 22.79)	5.14
Perception	(–8.4, –35.38, –6.21)	0	(3.7, –10.5, 27.49)	4.57
			(4.8, 0.13, 21.75)	4.91
			(4.9, 1.25, 22.31)	6.10
			(3.8, –6.38, 24.71)	1.07
			(5.0, –0.38, 22.79)	4.64

Q3: What is a support vector machine? What is the core?

Introduction of AI (CS103)- 08 Machine Learning Algorithms

Jimmy Liu 刘江

2023-10-10